DOCUMENT RESUME

ED 178 492 SP 014 828

AUTHOR Schutz, Robert W.; Gessaroli, Marc E.
TITLE The Effects of Block Size and Heterogeneity of Covariance on Type 1 Error Rates with Constant Error and Variable Error Data.
PUB DATE Jun 79
NOTE 21p.; Paper presented at the International Congress in Physical Education (Three Rivers, Quebec, Canada, June 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Analysis of Covariance; *Interaction; *Statistical Data; *Validity
IDENTIFIERS Block Size; Type 1 Error

ABSTRACT

The question of whether or not block size has an effect on the statistical analysis of data is examined through a study involving a total of 10,000 simulated experiments. Results indicate that the inflation in Type 1 error rate is much higher than is reported in the literature when using CE as the independent variable. There is little or no difference under varying blocking parameters. With VE, however, there was no effect of heterogeneity of covariance. It is concluded that block size can be chosen on logical rather than on statistical grounds, and that it has no effect on the validity of the statistical analysis involved. (Author/IH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED178492

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Robert W. Schutz
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Effects of Block Size and Heterogeneity of Covariance on Type 1 Error Rates with Constant Error and Variable Error Data

Robert W. Schutz, Marc E. Gessaroli

University of British Columbia

Paper presented at
International Congress in Physical Education
Three Rivers, Quebec
June, 1979

SP 014828

Running head: Heterogeneity of Covariance, Blocking, and Type 1 Error

Abstract

The problems addressed in this paper are: Does block size (number of trials per block) have an effect on the statistical analysis of this data? and, is the effect of heterogeneity of covariance reduced by increasing block sizes? Monte Carlo procedures were used to investigate the problem. Populations of 5000 subjects, with each subject having 36 trials reflecting a typical asymptotic learning curve, were generatef for variance-covariance matrices representing conditions of homogeneity and heterogeneity of variance. For each population 125 "experiments" were conducted for each of four different blocking sizes. An experiment consisted of randomly selecting 20 subjects for each of two groups, blocking the trials according to the specified block size, and then analyzing the data (both VE and CE) with a repeated measures ANOVA. The complete process was replicated three times, thus providing a total over 10,000 simulated experiments. The effect of interest was the Groups x Blocks interaction. Results indicated that the inflation in the Type 1 error rate is much higher than is reported in the literature when using CE as the dependent variable. There was no difference under different blocking parameters. With VE however, there was no effect of heterogeneity of covariance. It is concluded that block size can be chosen on logical, not statistical grounds, and has no effect on the validity of the statistical analysis.

The Effect of Block Size and Heterogeneity of Covariance on Type I Error Rates with Constant Error and Variable Error Data

Researchers employing repeated measures designs in motor learning studies are frequently confronted with two problems with respect to the analysis of their data. The first problem concerns heterogeneity of covariance and its effect on the probability of committing a type I error. As data from most motor learning experiments shows high adjacent trial correlations and low correlations for trials far apart, severe heterogeneity of covariance exists, with an associated highly inflated type I error rate. The careful researcher will account for this by utilizing multivariate techniques (Davidson, 1972; Finn, 1969), or adopting correction procedures as advocated by Box (1954b) or Greenhouse and Geisser (1959). On occasion however, the assumption of equality of covariances is not even tested and conventional analysis of variance procedures are applied. The second "problem", one not considered to be a problem by most investigators, concerns the choice of the number of trials per block. For example, in an experiment which has 60 trials the investigator may reduce his data to 10 blocks of 6 trials per block, or 6 blocks of 10 trials per block, or any of a number of other combinations. The choice of block size (number of trials per block) seems to be made on either aesthetic or random

criteria, not statistical ones. The researcher chooses a block size which will yield a relatively smooth curve, while still depicting the nature of any changes occurring over the learning phase. It appears that these two problems have been assumed to be quite unrelated. This assumption may in fact be false, as there is suggestive evidence that the effect of heterogeneity of covariance may be affected by block size. The purpose of this paper is to examine the hypothesis that the degree of inflation of type I error rates brought about by unequal covariances is inversely related to block size. The rationale for this hypothesis follows.

Collier, Baker, Mandeville, and Hayes (1969) have shown that in a repeated measures design with four measures per subject unequal covariances (values ranging from .80 to .20) cause an actual probability of committing a type I error (alpha) of .08 when using a nominal alpha level of .05. Furthermore, studies have shown the degree of inflation in the alpha level brought about by unequal variances in a randomized groups design is directly related to the number of groups (Box, 1954a, Glass, Peckham, & Sanders, 1972). Thus we know that heterogeneity of covariance (in a repeated measures design) inflates the type I error rate, and that the greater number of sets of data that contribute to the heterogeneity of variance (in a randomized groups design) the greater the inflation. However, there is no evidence

to show that an increase in the number of trials which contribute to the heterogeneity of covariance in a repeated measures design increases the degree of inflation of the actual alpha level. It is speculated here that the greater the number of trials, the greater the degree of inflation. If this is in fact true, then large block sizes, resulting in only a few block means being subjected to statistical analysis, result in minimal inflation of type I error rates. Small block sizes, resulting an analysis of a large number of block means, will cause a relatively large inflation in the actual alpha.

Method

Monte Carlo procedures were used to simulate 2 x 36 factorial experiments, repeated measures on the second factor, with 20 subjects in each of the two groups ("experimental" and "control"). One hundred twenty-five such experiments were simulated for each of a number of specified variance-covariance matrices utilizing the following procedures. Data, based upon a normal population, was produced for 36 trials for each of 5000 subjects. The data reflected an asymptotic learning curve, with either a homogeneous covariance structure (all intertrial correlations equal to .8) or a heterogenous covariance structure (correlations ranging from .90 to .50). The trial scores for each subject were then blocked into five different block sizes: (1) no blocks, (2) 12 blocks of 3 trials per block, (3) 9 blocks of 4 trials per block, (4) 6 blocks of 6 trials per block, (5) 4 blocks of 9 trials per

block. Constant error (CE) and variable error (VE) scores were calculated for each subject for each block. This resulted in five populations of 5000 subjects each, differing from each other only in the number of trials per block (and therefore the number of blocks). The final step assigned 20 subjects to group I, 20 to group II, and performed a two-way repeated measures analysis of variance on both CE and VE. The population of 5000 subjects thus provided data for 125 such experiments for each of the five blocking conditions. This complete procedure was performed twice for CE under both homogenous and heterogenous covariance conditions (i.e., 250 experiments per covariance matrix), and three times for VE with a heterogenous covariance matrix. Thus, with 5 different blocking conditions for each experiment, the total number of simulated experiments was 2500 using CE as the dependent variable and 1875 using VE as the dependent variable.

Results and Discussion

Each analyses of variance produced an F ratio and p value for the three sources of variance; Groups main effect, Blocks main effect, and the Groups by Blocks interaction. The Groups by Blocks effect was the only one examined as it is the source of variation of primary interest in most experiments of this design. As all subjects were sampled from a common population, the null hypothesis of no Groups by Blocks interaction was in fact true, and one would

expect the proportion of significant effects to be equal to the nominal alpha level set.

Homogeneity of Covariance

Table 1 shows that for the no blocking condition, and with homogenous among trial covariances, there was no difference between the percentage of significant results and the theoretical levels of significance. This, of course, was expected and merely supports the validity of the random generation and statistical procedures. Table 1 also reveals that this was true regardless of the blocking parameters. Thus it can be concluded that when the assumptions of equal variances and covariances are met the choice of block size can be made on non-statistical grounds.

Insert Table 1 about here

Heterogeneity of Covariance - CE

With unequal covariances and no blocking the experimental type 1 error rate was highly inflated, with 16% of false rejections for a nominal alpha level of .10, 12% with $\alpha = .05$, and 5% with $\alpha = .01$ (See Table 2). These obtained values are definitely significantly different from the nominal alpha values. With 250 simulations the standard error of a proportion for $\alpha = .10$ is .019, thus the obtained value of .16 is more than three standard errors in deviation - a highly unlikely chance occurrence. For $\alpha = .01$

the standard error is .006, thus the obtained value of .05 is more than six standard errors in deviation. These actual error rates were somewhat higher than expected and considerably larger than reported in the literature. Collier et al. (1967) have reported actual error rates of around .12 for $\alpha = .10$ and .03 for $\alpha = .01$, when using only four repeated measures and unequal covariances. It is speculated that the large number of repeated measures in the study (36) caused the increase in the inflation of type I error rates.

Insert Table 2 about here

It was hypothesized that by increasing the block size, and thereby decreasing the number of repeated measures per subject in the statistical analyses, the effect of heterogeneity of covariance would be reduced. This hypothesis was not supported by the data. As seen in Table 2 the degree of inflation in type I errors was constant across all blocking conditions. The assumption that the Mean Square error (Subjects within Groups by Blocks) would decrease with an increase in block size at a faster rate than would the Groups by Blocks Mean Square was in fact upheld. The F ratio increased from 1.034 for the no blocking condition up to 1.186 for the 4 blocks of 9 trials/block condition (See Table 3). However, the recuction in the degrees of freedom from both the numerator and

denominator of the F ratio caused an increase in the critical F which was sufficient to negate this increase in the F ratio. Consequently the proportion of false rejections was constant across all block conditions

---

Insert Table 3 about here

On the basis of the above results it must be concluded that researchers, when using CE as a dependent variable can choose blocking parameters on a completely arbitrary basis without prejudicing the validity of subsequent statistical analyses. Choosing a block size which provides maximum smoothing of data while at the same time providing enough points for an accurate description of the nature and degree change over trials cannot be challenged on statistical grounds.

Heterogeity of Covariance - VE

When VE data was studied unequal covariances appeared to have little or no effect upon the type I error rates. The percentage of significant results calculated at each level of significance did not differ from the nominal alpha levels. As seen in Table 2, all of the obtained significant findings were well within two standard errors of the theoretical level of significance for the no blocking condition. These results were consistent across all blocking conditions for the alpha levels of .10 and .05, but not for .01.

The obtained values of .021 and .029, for the nominal alpha level of .01, at the 9 x 4 and 6 x 6 blocking conditions suggest some inflation. These discrepant findings were noticed on completion of 250 simulations, and as there seemed to be no plausible explanation why such an inflated alpha level would occur only at the .01 level, and only under these two specific blocking conditions, another 125 simulated experiments were run. Analysis of these additional ANOVAS only confirmed the previous results. It is not clear why this occurred - obviously further work is needed in this area.

A logical conclusion of these findings might be that heterogeneity of covariance does not inflate type 1 error rates when VE is the dependent variable. Thus published studies which make inferences about the organization and execution of motor programming on the basis of significant CE results and nonsignificant VE results may be quite misleading. It is feasible that the significant CE results were due to a type 1 error brought about by heterogeneity of covariance - a condition which more than doubles the assumed probability for making a type 1 error with CE data, but does not inflate the probability for VE data. The above conclusion is somewhat tenuous and could be erroneous due to an assumption under which this data was simulated. The data for subjects was highly correlated across trials, thus reflecting true intraindividual differences in raw scores, and therefore CE scores. This is congruent with actual experimental data. The intraindividual variability however was

not controlled, and thus VE scores were uncorrelated across blocks. As it is not unusual for empirical data to exhibit correlated VE scores, the simulated data, exhibiting uncorrelated VE scores, is not realistic. This in itself is not a problem, as the mathematical assumption of zero covariances is not necessary for unbiased analysis. What is important is the equality of the covariances, whatever magnitude they may be. Thus if actual experimental data exhibits among block covariances which are large, but equal, the actual probability of committing a type 1 error is equal to the nominal alpha value and the conclusions of this study are valid. If, on the other hand, empirical data show covariances with VE data which tend to decrease as the distance between blocks increases (as is the case with CE data), then the conclusion of this study are restrictive.

A thorough examination of the presence or absence of heterogeneity of covariance in motor learning data has not yet been conducted. Two sets of convenient data were examined however, and results suggest that even with heterogeneity of raw score data, and therefore of CE data, VE scores are relatively homogeneous. Data from a kinesthetic reproduction task involving 25 trials (Stafford, 1978) showed adjacent trial correlations of .50 to .72, with distant trial correlations of zero to slightly negative values. When blocked into 5 blocks of 5 trials per block, VE scores were corre-

lated approximately .20 for both adjacent blocks and distant blocks. Data from the Smoll and Schutz (1978) motor rhythm study involved 40 trials with temporal accuracy as the dependent measure. Adjacent trial correlations were approximately .85, with distant trials correlating between .50 and .60 (very similar to the generated data of this study). When blocked into 10 blocks of 5 trials/block, the adjacent block VE scores correlated at about .40, and distant blocks at approximately .30. Thus for both these data sets there was relatively large heterogeneity of variance in raw scores and blocked CE data, but the blocked VE data, although still correlated, exhibited lower and much more homogeneous correlations. On the basis of this small bit of evidence there is no reason to discount assumptions upon which the data for this study were simulated. Further investigations into the assumptions of homogeneity of covariance for VE data is needed.

The Solution

Researchers can do very little to reduce heterogeneity of covariance without destroying the validity of their experiment. The problem, and therefore the solution, lies with the statistical analysis, not the data. Although approximate adjustments can be made to compensate for the distortion in alpha (e.g., Greenhouse & Geisser, 1959), multivariate analysis of variance provides an exact solution (Davidson, 1972; Poor, 1973). MANOVA, which requires no assumptions of variance or covariance homogeneity because of the orthogonal decomposition of the repeated measures, has been frequently recommended as the appropriate technique for all repeated measures designs,

regardless of variance-covariance homogeneity (e.g., see Schutz, 1978). For those investigators who do not have access to an appropriate MANOVA computer program, an ANOVA with trend analysis (or any other orthogonal breakdown) on the trials factor will do just as well. An example of this is provided in Table 4. One set of experiments (125 ANOVAs) was analyzed using a trend analyses on the blocks factors from the CE data with heterogeneity of covariance. As Table 4 indicates, the Groups by Blocks term has highly inflated type 1 error rates, but the trend components of interest to most experimenters, the Groups by Blocks (linear) and Groups by Blocks (quadratic) show no such inflation. Clearly this is an appropriate analysis.

Conclusions

On the basis of the Monte Carlo procedures reported here, the following conclusions can be postulated:

i) Analyses of variance procedures conducted on data exhibiting a large number of trials (36) with moderate heterogeneity of covariance (correlations from .80 to .50) result in highly inflated type 1 error rates on the Groups by Blocks interaction significance test. Actual error rates for the nominal alpha rates of .10, .05 and .01, are .17, .12, and .05 respectively.

ii) Grouping the trials into blocks, and then using the CE of each block as the dependent measure, does not affect the magnitude of the type 1 error inflation rate. This holds true regardless of block size.

iii) Grouping the trials into blocks, and then using VE of each block as the dependent measure, eliminates the heterogeneity of covariance and consequently the actual type 1 error rate is congruent with the nominal rate. Varying the block size may introduce some inflation under certain conditions, although this is not clear from the data of the study and further research is required.

iv) Because heterogeneity of covariance results in a highly inflated type 1 error rate for CE, but not for VE, one must be cautious in interpreting some of the published literature. Studies which report significant VE effects, and then state conclusions regarding the organization and execution of motor programs on the basis of these findings, fall into this category. The inferences could be valid - or a type 1 error could have been made with CE - one cannot tell which is the case.

v) Multivariate analysis of vairance, or univariate analysis with an orthogonal breakdown on the blocks factor, is not based on the assumption of variance-covariance homogeneity and thus is the appropriate statistical analysis for repeated measures designs.

TABLE 1

Proportion of Significant Groups x Trials Interaction:

Homogeneity of Covariance, CE Data

Block Size	Nominal Alpha	Proportion Significant
No Blocking	.10	.094
	.05	.048
	.01	.008
12 Blocks of 3 Trials	.10	.100
	.05	.064
	.01	.008
9 Blocks of 4	.10	.112
	.05	.056
	.01	.008
6 Blocks of 6	.10	.108
	.05	.044
	.01	.000
4 Blocks of 9	.10	.108
	.05	.068
	.01	.008

TABLE 2

Proportion Significant of Group x Trials Interaction:
Heterogeneity of Covariance, CE and VE Data

Block Size	CE			VE		
	.10	.05	.01	.10	.05	.01
No blocks	.16	.12	.05			
12 blocks of 3 trials/block	.17	.13	.06	.11	.06	.01
9 blocks of 4 trials/block	.17	.12	.04	.09	.04	.02
6 blocks of 6 trials/block	.17	.12	.06	.12	.06	.03
4 blocks of 9 trials/block	.18	.10	.05	.10	.05	.01
Averaged across blocking conditions	.169	.116	.051	.103	.055	.018
Standard error of proportion *	.019	.014	.006	.015	.011	.005

* The CE values were based on 250 simulations, the VE on 375, thus the different values for the standard error of a proportion.

TABLE 3

Block Size and Magnitudes of Mean Squares:
Heterogeneity of Covariance, CE Data

	Average Mean Squares and F Ratios*				
Source of Variation	No Blocking	12 Blocks of 3	9 Blocks of 4	6 Blocks of 6	4 Blocks of 9
Groups x Blocks	.333	.169	.149	.129	.118
SwG x B	.322	.155	.134	.113	.099
F	1.034	1.090	1.112	1.142	1.186
df For F	35,1330	11,418	8,304	5,190	3,114
Critical F at .10	1.32	1.61	1.69	1.89	2.13
Number Sig. at $\leq$.10	22	23	23	24	25

*These values are based on only 125 of the 250 simulations

TABLE 4

Proportion of Significant CE Effects when Analyzed by a Multivariate and/or Orthogonal Breakdown on Trials

	4 Blocks of 9			6 Blocks of 6		
	.10	.05	.01	.10	.05	.01
G x B	.18	.10	.05	.17	.12	.06
G x B (Lin)	.12	.06	.016	.13	.07	.016
G x B (Quad)	.09	.05	.016	.11	.05	.016

References

Box, G.E.P. Some theorems on quadratic forms applied in the study of analysis of variance problems, I: Effect of inequality of variance in the one-way classification. Annals of Mathematical Statistics, 1954, 25, 290-302. (a)

Box, G.E.P. Some theorems on quadratic forms applied in the study of variance problems, II: Effects of inequality of variance and of correlation between errors in the two-way classification. Annals of Mathematical Statistics, 1954, 25, 484-498. (b)

Collier, R.O., Baker, F.B., Mandeville, G.K., & Haynes, T.F. Estimation of test size for several test procedures based on conventional variance ratios in the repeated measures design. Psychometrika, 1967, 32, 339-353.

Davidson, M.L. Univariate versus multivariate tests in repeated measures experiments. Psychological Bulletin, 1972, 77, 446-452.

Finn, J.D. Multivariate analysis of repeated measures data. Multivariate Behavioural Research, 1969, 4, 391-413.

Glass, G.V., Peckham, P.D., & Sanders, J.R. Consequences of Failure to Meet Assumptions Underlying The Fixed Effects Analyses of Variance and Covariance. Review of Educational Research, 1972, 42, 237-288.

Poor, D.D.S. Analysis of variance for repeated measures designs: Two approaches. Psychological Bulletin, 1973, 80, 204-209.

Smoll, F.L. and Schutz, R.W. Relationships Among Measures of Preferred Tempo and Motor Rhythm. Perceptual and Motor Skills, 1978, 46, 883-894.

References (cont'd)

Schutz, R.W. Specific Problems in the Measurement of Change: Longitudinal Studies, Difference Scores, and Multivariate Analyses, in Landers, D. and Christina, R. (Eds.) Psychology of Motor Behavior & Sport - 1977. Champaign, Illinois. Human Kinetics, 151-175, 1978.

Stafford, E.M. Knowledge of results and the perceptual trace. Unpublished Master's Thesis. University of British Columbia, 1978.